Ph.D. Thesis Synopsis

Interplay between enhancer hierarchies and histone demethylase JMJD3

Submitted to
The University of Trans-Disciplinary Health Sciences and Technology (TDU) - Bengaluru - 560064

By

Umer Farooq
(21117020496)

In partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Thesis Supervisor: **Dr. Dimple Notani**

Introduction

Enhancers in transcription regulation

Cell differentiation, development, responses to external stimuli, and other processes rely on precise and accurate gene expression, which is regulated at various stages by several protein factors, RNAs, and DNA regulatory elements (Buccitelli and Selbach 2020; Pope and Medzhitov 2018). The key DNA regulators of transcription are cis-regulatory DNA elements called enhancers. They are short DNA elements ranging from 100 to 1500 bp and induce target gene transcription regardless of their proximity, orientation, or placement with respect to the gene promoter (Panigrahi and O'Malley 2021; Banerji, Rusconi, and Schaffner 1981). Enhancers are cell type-specific, meaning a specific set of enhancers that are active in one cell type may be inactive in another. Certain modifications, such as H3K4me1 and H3K27ac, are found on the histones of active enhancers (Li, Notani, and Rosenfeld 2016). Active enhancers exhibit transcription factor and cofactor binding and are hypersensitive to DNase I treatment due to their core accessibility. Furthermore, because of their ability to recruit PolII, active enhancers are transcribed into non-coding RNAs known as enhancer RNAs (eRNAs) (Li, Notani, and Rosenfeld 2016). These non-coding RNAs have been shown to increase target promoter transcription by enhancing enhancer activity (Arnold, Wells, and Li 2019; Li et al. 2013). Enhancer-mediated gene transcription regulation is thought to occur via looping between the enhancer and the target gene promoter (Panigrahi and O'Malley 2021). Although enhancer-promoter (E-P) looping is a widely accepted model of gene regulation, the underlying mechanism is unclear.

Enhancers can appear in clusters composed of multiple individual enhancers that coexist in close linear proximity. These clustered enhancers are called super-enhancers if the enhancers within the cluster occur in close linear proximity to one another and exhibit exceptionally high enrichment of enhancer-associated marks (Hnisz et al. 2013). These enhancers can phase separate by liquid-liquid phase-separated (LLPS) due to the abundance of transcription factors and cofactors such as Mediator (Sabari et al. 2018; Whyte et al. 2013). Super-enhancers are primarily responsible for driving the expression of genes important for cell identity (Hnisz et al. 2013). These enhancers, for example, have been discovered near pluripotency genes such as Sox2, Nanog, and others in mESCs (Li et

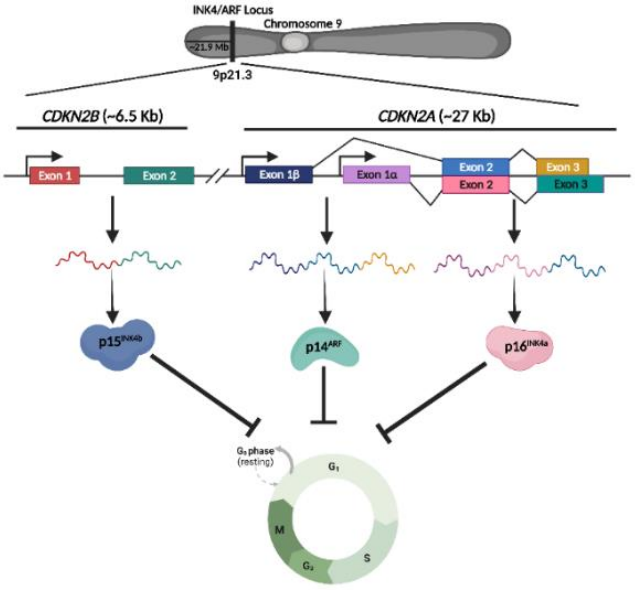
al. 2014; Blinka et al. 2016). Furthermore, super-enhancers are abundant around oncogenes in cancer cells and have been shown to increase the expression of these genes dramatically. For instance, they have been found in the MYC locus in several cancers (Schuijers et al. 2018). Thus, understanding how super-enhancers work holds enormous therapeutic potential.

The thesis aimed to understand how dense enhancer clusters function. Do all the individual enhancers that form the cluster regulate target gene transcription or only a subset is important? What is the relationship between functional enhancers? Thus, to dissect the individual regulatory enhancer elements from the dense enhancer cluster and to understand the intrinsic properties of these regulatory elements, we chose INK4/ARF locus. This locus harbors a very dense enhancer cluster. Moreover, the enhancer cluster is already associated with several aging-associated diseases like coronary artery disease (CAD), type 2 diabetes, and several cancers.

INK4/ARF locus:

The INK4/ARF locus contains two crucial cell cycle regulatory genes: *CDKN2A*, which codes for $p14^{ARF}$ and $p16^{INK4a}$, and *CDKN2B*, which codes for $p15^{INK4b}$. Another gene at this locus, *CDKN2BAS*, codes for the lncRNA ANRIL. Although the mRNA sequences of $p14^{ARF}$ and $p16^{INK4a}$ are very similar, the resulting proteins differ due to different reading frames. However, the amino acid sequences of $p16^{INK4a}$ and $p15^{INK4b}$ encoded by different genes are very similar. Thus, gene duplication events are thought to have resulted in the *CDKN2A* and *CDKN2B* genes (López et al. 2017). Despite their differences, all three proteins play cell cycle regulatory roles and cause cell cycle arrest, primarily in the G1 phase. In normal cells, these genes are suppressed, but they become active as the cell ages and enters senescence. Furthermore, this locus is either deleted or silenced by repressive machinery in most cancers. These proteins, however, are overexpressed in a few cancers, including head and neck carcinoma and HPV-expressing cancers like cervical cancer. This is made possible by two viral proteins, E6 and E7. Because E6 induces degradation of p53

and E7 inhibits Rb, these proteins cannot activate their downstream targets which result in cell cycle inhibition (Tommasino and Crawford 1995).



Schematic representing the INK4/ARF locus. *This locus contains two critical cell cycle regulatory coding genes: CDKN2A and CKDN2B. CDKN2A encodes two proteins, p14ARF and p16INK4a, whereas CDKN2B encodes p15INK4b. Though distinct, all three proteins perform cell cycle regulatory functions and arrest the cell cycle at the G1 phase.*

Regulation of INK4/ARF locus via distal regulatory elements

A long gene desert region exists upstream of the INK4/ARF locus, and SNPs in this region have been linked to various diseases, such as coronary artery disease, type 2 diabetes, and multiple cancers. In mice, deletion of this syntenic CAD interval reduced the expression of these genes significantly (Visel et al. 2010). A seminal study discovered several enhancers present in the gene desert to understand the functional relevance of gene desert in regulating this locus. Furthermore, SNP in one of the enhancers disrupts the STAT1 motif, causing the genes to be downregulated in response to IFNγ signaling (Harismendy et al. 2011). TGFβ is required to activate upstream enhancers and binds to one of the CAD interval enhancers, promoting ARF expression during development. TGFβ also induces three H3K27ac peaks in the gene desert, and deletion of the entire interval reduces INK4b and

ARF expression (Zheng et al. 2013; Liu et al. 2019). Furthermore, people with the SNP rs10757278 have significantly lower expression of all three genes in T cells in their peripheral blood. A cis-regulatory element near the ARF promoter, in addition to upstream regulatory elements, inhibits INK4a gene expression by forming a loop with its promoter (Zhang, Hyle, et al. 2019). All studies link gene desert enhancers to INK4/ARF locus transcription regulation. However, how multiple enhancers in this gene desert activate the locus was unknown. My work towards the doctoral thesis has revealed an intricate relationship among functional enhancers in this gene desert.

Regulation of INK4/ARF locus via chromatin modifiers

1. **By Polycomb Complexes:**

Polycomb group proteins (PcGs) are methyltransferase enzyme complexes necessary for development, tissue repair, differentiation, aging, and other processes. PcGs form two different multi-subunit complexes: PRC1 and PRC2 (Chittock et al. 2017). PRC1 is comprised of several core subunits, including RING1A/B, PCGF2/4, CBX2/4/6/7/8, and PHC1/2/3. In contrast, the PRC2 complex is comprised of core subunits such as SUZ12, EZH2, RbAp46/48, and EED (Chittock et al. 2017; Kerppola 2009). In addition to core subunits, other auxiliary subunits have been discovered that aid the function of both these complexes (Kerppola 2009). Both PRC1 and PRC2 complexes are required for INK4/ARF locus suppression in proliferating cells. PRC2 accomplishes this by binding to and trimethylating the lysine 27 of histone 3 (H3K27me3) at promoters of INK4/ARF genes (Bracken et al. 2007). Ectopic expression of PRC1 and PRC2 subunits suppresses this locus and delays aging (Dietrich et al. 2007). On the other hand, downregulation of the PcG subunits results in the activation of this locus and premature senescence.

2. **By JMJD3:**

JMJD3 is a histone demethylase essential for development, infectious disease progression, cancer progression, senescence, etc. (Xiang et al. 2007; Zhang, Liu, et al. 2019). It acts as

an antagonist to the PRC2 complex by removing methyl marks from Histone 3 trimethylated at lysine 27 (H3K27me3). JMJD3 is a well-known positive regulator of the INK4/ARF locus and during senescence, it activates this locus by removing the repressive methylation (Barradas et al. 2009). JMJD3 is activated by cellular signals such as oncogene overexpression which activates the INK4/ARF locus (Agger et al. 2009). As a result, INK4/ARF-mediated cell cycle arrest occurs in MEFs and IMR90 cells. Furthermore, JMJD3 acts as a negative regulator for reprogramming by activating the INK4/ARF locus (Zhao et al. 2013). Silencing JMJD3 alone increases reprogramming efficiency by orders of magnitude. Surprisingly, inhibiting JMJD3 with INK4a or ARF improves reprogramming efficiency even further (Zhao et al. 2013).

The study found that only a few enhancers that looped with the promoter were important and they had an intricate mutual relationship. Later, the JMJD3 (a histone demethylase) was functionally tied with the enhancer functions and hierarchy. A brief outline of thesis chapters that include, the description of the INK4/ARF locus, its regulation, results of the study, material and methods used, interpretations, references, and publications are mentioned below.

Thesis Organization

Chapter 1 (Introduction) of the thesis discusses the existing literature on DNA regulatory elements called enhancers and their role in transcription regulation. It also introduces other aspects of enhancers, including active enhancer signatures, methods and approaches for identifying regulatory enhancers, and the role of enhancers in higher-order chromatin organization. Furthermore, it also summarizes the effects of SNPs in regulatory enhancers and their impact on target gene transcription and subsequent association with several diseases. This chapter also introduces INK4/ARF locus and its role in cell cycle regulation. Furthermore, it also explains the role of this locus in cancers, reprogramming, and senescence. Moreover, this chapter also focuses on the known cis and trans mechanisms of transcription regulation of this locus.

Chapter 2 (Findings-Part I) explains how individual functional enhancers that regulate the INK4/ARF locus were dissected from a dense enhancer cluster. The first section of this chapter discusses the identification of an enhancer cluster upstream of the INK4/ARF locus and its relationship to *CDKN2A* and *CDKN2B* gene expression. This enhancer cluster was found to contain 24 individual enhancers in HeLa, only 15 of which are active. The active enhancers also correlate with the expression of INK4/ARF genes in this cell line. The second section discusses the dissection of regulatory enhancers from this dense enhancer cluster. Only 5 of the 15 active enhancers were found to be interacting with the promoters. CRISPR-Cas9 perturbations were used to test the regulatory potential of promoter-interacting enhancers. It was found that promoter-interacting enhancers are required for INK4/ARF transcription, and they were found to be interdependent on one another. The third section discusses the impact of perturbation of these enhancers on the cancer phenotype of the cells. The deletion of the promoter-interacting enhancers affects the cancerous properties of the cells, such as the colony-forming potential, cell migration, etc. The final section discusses the mechanism of gene expression downregulation in the absence of enhancers. When promoter-interacting enhancers are disrupted, EZH2, a component of the repressive complex, is loaded onto the promoters of these genes, explaining why transcription is downregulated in the absence of promoters. Furthermore, the genes that are downregulated when the enhancers are deleted corroborate with the disease association of this locus.

Chapter 3 (Findings-Part II) shows that JMJD3 binding to INK4/ARF enhancers and other genome-wide regions depends on its N-terminal region containing an intrinsically disordered region (IDR) and the RNA binding domain. The first part discusses the genome-wide binding of JMJD3 on chromatin. It was observed that JMJD3 binds more to enhancer elements than promoters across the genome. Its binding pattern on regulatory elements differs from that of UTX, which binds more to promoters than enhancers. The second section explains how JMJD3's chromatin binding depends on its IDR. When the IDR of JMJD3 is removed, it no longer binds to chromatin. However, removing half of its IDR alters its binding pattern across the genome. The third section discusses JMJD3's chromatin binding dependence on RNA. We found that JMJD3 is an RNA-binding protein and using

domain mapping, it was found that its C-terminus, a Zinc finger-like domain, binds to RNA. Furthermore, JMJD3 chromatin targeting is RNA-dependent, and it binds to the INK4/ARF locus enhancers in an eRNA-dependent manner. The final section discusses the effects of removing its IDR region and mutating its RNA binding on its catalytic activity. Removing half of JMJD3's IDR does not affect its catalytic activity, but removing the entire IDR reduces its catalytic activity. Moreover, RNA-binding mutants also show reduced catalytic activity.

Chapter 4 (Materials and Methods) describes the materials and methods used in this study. To address the questions of Chapter 2 (Finding Part-I), we took advantage of techniques and approaches like Chromatin Immunoprecipitation (ChIP), CRISPR-Cas9 assays, Chromosome conformation capture techniques (4C), Gene expression, etc. To address the questions of Chapter 3 (Finding Part II), we utilized approaches like ChIP-seq, RNA Immunoprecipitations (UV-RIP), RNA Pulldown Assays, Site-directed mutagenesis, and Subcellular fractionation.

Chapter 5 (Conclusions and Discussion) summarizes the key findings from chapter 2 and chapter 3. Our results in chapter 2 highlight that only a subset of enhancers from a dense enhancer cluster can regulate the target genes. The enhancers interacting with the promoters are critical for the transcriptional regulation of the INK4/ARF locus. The promoter-interacting enhancers are mutually dependent on each other for function. The perturbation of one enhancer in the cluster affects the other enhancers in the same cluster. The deletion of promoter-interacting enhancers affects the cancerous phenotype of the cells by reducing their colony-forming potential and delaying migration confirming, the role of this locus in the cell cycle and cancer. When regulatory enhancers are deleted, EZH2 is loaded onto INK4/ARF, causing the genes to be down-regulated. Furthermore, the dysregulated genes after enhancer deletion correlate with the diseases this locus is associated with. Our findings in Chapter 3 show that JMJD3 binds to enhancers more than promoters. The binding of JMJD3 on chromatin is dependent on its N-terminal IDR, the complete removal of which affects its chromatin binding whereas, the removal of half IDR does not affect the chromatin binding but its target specificity is altered. These results

suggest that the IDR of JMJD3 is essential for chromatin binding and determining target specificity. Furthermore, JMJD3 chromatin binding is RNA-dependent, and mutations in its RNA binding domain affect chromatin binding.

Publications:

Farooq, U, Saravanan, B., Islam, Z., Walavalkar, K., Singh, A.K., Jayani, R.S., Meel, S., Swaminathan, S., Notani, D. An interdependent network of functional enhancers regulate transcription and EZH2 loading at INK4a/ARF locus. *Cell Reports*, 34 (12), 108898. (2021)

Farooq, U and Notani, D. Optimized protocol to create deletion in adherent cell lines using CRISPR/Cas9 system. *STAR Protocols*, 2(4), 100857. (2021)

Farooq U* and Notani D*, Transcriptional regulation of INK4/ARF locus by cis and trans mechanisms. *Front. Cell Dev. Biol.* 10:948351. (2022) *Corresponding author

Walavalkar, K., Saravanan, B., Singh, A.K., Jayani, R.S., Nair, A., **Farooq, U.,** Islam, Z., Soota, D., Mann, R., Padubidri, S.V., Freedman, M.L. Sabarinathan. R., Haiman, C., Notani, D. A rare variant of African ancestry activates 8q24 lncRNA hub by modulating a cancer. *Nature communications*, 11, 3598. (2020)

Saravanan, B., Soota, D., Islam, Z., Majumdar, S., Mann, R., Meel, S., **Farooq, U.,** Walavalkar, K., Gayen, S., Singh, A.K., Hannenhalli, S. and Notani, D. Ligand dependent gene regulation by transient ERα clustered enhancers. *PLoS genetics*, 16(1), p.e1008516. (2020)

Islam, Z.*, Saravanan, B.*, Walavalkar, K., **Farooq, U.,** Singh, A.K., Thakur, J., Pandit, A., Sabarinathan, R., Henikoff, S., Notani, D. Active enhancers strengthen TAD insulation by bRNA mediated CTCF enrichment at the TAD boundaries. *(Genome Research, In revision*)

References:

Agger, K., P. A. Cloos, L. Rudkjaer, K. Williams, G. Andersen, J. Christensen, and K. Helin. 2009. 'The H3K27me3 demethylase JMJD3 contributes to the activation of the INK4A-ARF locus in response to oncogene- and stress-induced senescence', *Genes Dev*, 23: 1171-6.

Arnold, P. R., A. D. Wells, and X. C. Li. 2019. 'Diversity and Emerging Roles of Enhancer RNA in Regulation of Gene Expression and Cell Fate', *Front Cell Dev Biol*, 7: 377.

Banerji, J., S. Rusconi, and W. Schaffner. 1981. 'Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences', *Cell*, 27: 299-308.

Barradas, M., E. Anderton, J. C. Acosta, S. Li, A. Banito, M. Rodriguez-Niedenfuhr, G. Maertens, M. Banck, M. M. Zhou, M. J. Walsh, G. Peters, and J. Gil. 2009. 'Histone demethylase JMJD3 contributes to epigenetic control of INK4a/ARF by oncogenic RAS', *Genes Dev*, 23: 1177-82.

Blinka, Steven, Michael H Reimer, Kirthi Pulakanti, and Sridhar Rao. 2016. 'Super-Enhancers at the Nanog Locus Differentially Regulate Neighboring Pluripotency-Associated Genes', *Cell Reports*, 17: 19-28.

Bracken, A. P., D. Kleine-Kohlbrecher, N. Dietrich, D. Pasini, G. Gargiulo, C. Beekman, K. Theilgaard-Monch, S. Minucci, B. T. Porse, J. C. Marine, K. H. Hansen, and K. Helin. 2007. 'The Polycomb group proteins bind throughout the INK4A-ARF locus and are disassociated in senescent cells', *Genes Dev*, 21: 525-30.

Buccitelli, C., and M. Selbach. 2020. 'mRNAs, proteins and the emerging principles of gene expression control', *Nat Rev Genet*, 21: 630-44.

Chittock, E. C., S. Latwiel, T. C. Miller, and C. W. Muller. 2017. 'Molecular architecture of polycomb repressive complexes', *Biochem Soc Trans*, 45: 193-205.

Dietrich, N., A. P. Bracken, E. Trinh, C. K. Schjerling, H. Koseki, J. Rappsilber, K. Helin, and K. H. Hansen. 2007. 'Bypass of senescence by the polycomb group protein CBX8 through direct binding to the INK4A-ARF locus', *EMBO J*, 26: 1637-48.

Harismendy, O., D. Notani, X. Song, N. G. Rahim, B. Tanasa, N. Heintzman, B. Ren, X. D. Fu, E. J. Topol, M. G. Rosenfeld, and K. A. Frazer. 2011. '9p21 DNA variants

associated with coronary artery disease impair interferon-gamma signalling response', *Nature*, 470: 264-8.

Hnisz, D., B. J. Abraham, T. I. Lee, A. Lau, V. Saint-André, A. A. Sigova, H. A. Hoke, and R. A. Young. 2013. 'Super-enhancers in the control of cell identity and disease', *Cell*, 155: 934-47.

Kerppola, T. K. 2009. 'Polycomb group complexes--many combinations, many functions', *Trends Cell Biol*, 19: 692-704.

Li, W., D. Notani, Q. Ma, B. Tanasa, E. Nunez, A. Y. Chen, D. Merkurjev, J. Zhang, K. Ohgi, X. Song, S. Oh, H. S. Kim, C. K. Glass, and M. G. Rosenfeld. 2013. 'Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation', *Nature*, 498: 516-20.

Li, Wenbo, Dimple Notani, and Michael G. Rosenfeld. 2016. 'Enhancers as non-coding RNA transcription units: recent insights and future perspectives', *Nature Reviews Genetics*, 17: 207-23.

Li, Yan, Chloe M. Rivera, Haruhiko Ishii, Fulai Jin, Siddarth Selvaraj, Ah Young Lee, Jesse R. Dixon, and Bing Ren. 2014. 'CRISPR Reveals a Distal Super-Enhancer Required for Sox2 Expression in Mouse Embryonic Stem Cells', *PLoS One*, 9: e114485.

Liu, Y. T., L. Xu, L. Bennett, J. C. Hooks, J. Liu, Q. Zhou, P. Liem, Y. Zheng, and S. X. Skapek. 2019. 'Identification of De Novo Enhancers Activated by TGFbeta to Drive Expression of CDKN2A and B in HeLa Cells', *Mol Cancer Res*, 17: 1854-66.

López, Fernando, Teresa Sampedro, José L. Llorente, Mario Hermsen, and César Álvarez-Marcos. 2017. 'Alterations of p14ARF, p15INK4b, and p16INK4aGenes in Primary Laryngeal Squamous Cell Carcinoma', *Pathology & Oncology Research*, 23: 63-71.

Panigrahi, A., and B. W. O'Malley. 2021. 'Mechanisms of enhancer action: the known and the unknown', *Genome Biol*, 22: 108.

Panigrahi, Anil, and Bert W. O'Malley. 2021. 'Mechanisms of enhancer action: the known and the unknown', *Genome Biology*, 22: 108.

Pope, S. D., and R. Medzhitov. 2018. 'Emerging Principles of Gene Expression Programs and Their Regulation', *Mol Cell*, 71: 389-97.

Sabari, B. R., A. Dall'Agnese, A. Boija, I. A. Klein, E. L. Coffey, K. Shrinivas, B. J. Abraham, N. M. Hannett, A. V. Zamudio, J. C. Manteiga, C. H. Li, Y. E. Guo, D. S. Day, J. Schuijers, E. Vasile, S. Malik, D. Hnisz, T. I. Lee, Cisse, II, R. G. Roeder, P. A. Sharp, A. K. Chakraborty, and R. A. Young. 2018. 'Coactivator condensation at super-enhancers links phase separation and gene control', *Science*, 361.

Schuijers, J., J. C. Manteiga, A. S. Weintraub, D. S. Day, A. V. Zamudio, D. Hnisz, T. I. Lee, and R. A. Young. 2018. 'Transcriptional Dysregulation of MYC Reveals Common Enhancer-Docking Mechanism', *Cell Rep*, 23: 349-60.

Tommasino, M., and L. Crawford. 1995. 'Human papillomavirus E6 and E7: proteins which deregulate the cell cycle', *Bioessays*, 17: 509-18.

Visel, A., Y. Zhu, D. May, V. Afzal, E. Gong, C. Attanasio, M. J. Blow, J. C. Cohen, E. M. Rubin, and L. A. Pennacchio. 2010. 'Targeted deletion of the 9p21 non-coding coronary artery disease risk interval in mice', *Nature*, 464: 409-12.

Whyte, W. A., D. A. Orlando, D. Hnisz, B. J. Abraham, C. Y. Lin, M. H. Kagey, P. B. Rahl, T. I. Lee, and R. A. Young. 2013. 'Master transcription factors and mediator establish super-enhancers at key cell identity genes', *Cell*, 153: 307-19.

Xiang, Y., Z. Zhu, G. Han, H. Lin, L. Xu, and C. D. Chen. 2007. 'JMJD3 is a histone H3K27 demethylase', *Cell Res*, 17: 850-7.

Zhang, X., L. Liu, X. Yuan, Y. Wei, and X. Wei. 2019. 'JMJD3 in the regulation of human diseases', *Protein Cell*, 10: 864-82.

Zhang, Y., J. Hyle, S. Wright, Y. Shao, X. Zhao, H. Zhang, and C. Li. 2019. 'A cis-element within the ARF locus mediates repression of p16 (INK4A) expression via long-range chromatin interactions', *Proc Natl Acad Sci U S A*.

Zhao, W., Q. Li, S. Ayers, Y. Gu, Z. Shi, Q. Zhu, Y. Chen, H. Y. Wang, and R. F. Wang. 2013. 'Jmjd3 inhibits reprogramming by upregulating expression of INK4a/Arf and targeting PHF20 for ubiquitination', *Cell*, 152: 1037-50.

Zheng, Y., C. Devitt, J. Liu, J. Mei, and S. X. Skapek. 2013. 'A distant, cis-acting enhancer drives induction of Arf by Tgfbeta in the developing eye', *Dev Biol*, 380: 49-57.